Measurement Uncertainty and Traceability Issues
in National and International Measurements

Presented at the International Dimensional Metrology Workshop, Nashville, TN
May 12-16, 2003

Dr. Steven D. Phillips
Precision Engineering Division
National Institute of Standards and Technology
Gaithersburg, Maryland 20899

Abstract

This paper discusses some recent laboratory intercomparisions with emphasis on the success of the uncertainty statement to include the reference value. Some factors that affect this capability are discussed. Recently developed national and international standards in the area of measurement uncertainty are presented as resources for industrial metrologists.

Introduction

The field of metrology has always been focused on measurement accuracy, which is defined as the closeness of agreement between the result of a measurement and the true value of the measurand. A decade ago the formalism for quantitatively expressing accuracy was published by the ISO in the "Guide to the Expression of Uncertainty in Measurement" (GUM) [1]. It is now the definitive document on evaluating measurement uncertainty. It is a remarkably self-consistent and complete document and has been adopted by National Measurement Institutes (NMIs), including NIST in the United States. In 1997 the GUM was adopted as a national (ANSI) standard in the US and is designated NCSL Z540-2-1997.

In recent years, considerable interest has developed over issues concerning measurement uncertainty and traceability. The motivations for this are many. The globalization of the economy allows industry to outsource workpiece production and inspection on a worldwide scale. Hence component interchangeability (not only between components produced by one supplier but also between the same nominal component produced by several different suppliers) can be assured only if all suppliers employ metrology to a common set of units (typically the SI units). Similarly, inspection services are frequently outsourced and the magnitude of the measurement uncertainty is taken as a measure of the quality and reliability of the measurement result; so measurement uncertainty is becoming a currency of metrology. Concomitantly, various national and international quality standards and laboratory accreditation programs are being revised to include language addressing measurement uncertainty and traceability. Finally, as workpiece tolerances steadily decrease, the cost of inspection usually increases, thus the ability to easily achieve a 10:1 ratio of the tolerance interval to measurement uncertainty interval is increasingly difficult or impossible. Measurement uncertainty is also affecting the economics of production through the cost of expensive equipment and facilities to perform metrology and in the cost of incorrect decisions, e.g., rejecting conforming workpieces or accepting nonconforming ones. Hence optimizing measurement uncertainty in an economic sense is now becoming an important issue in both laboratories and industry. This paper briefly reviews the capabilities of metrologists to create reasonable uncertainty statements and provides some recently developed standards and other documents as a resource helpful in creating uncertainty statements.

Measurement Uncertainty Statements

An expanded uncertainty statement is meant to encompass a large fraction of the values that can be reasonably attributed to the measurand. The concept of reasonableness inherently invokes judgment, prior information, as well as classical (frequentist) statistics. Consequently, it can be argued that there is no single "right answer" or "correct value" for the expanded uncertainty. It depends not only on the measurement system but also on the totality of the experience and knowledge of the metrologist. Hence, two different metrologists can perform two nearly identical measurements on the same measurand using the same measurement system and produce two quantitatively different uncertainty statements. Each uncertainty statement would be a personal statement of belief or "state of knowledge" concerning what can be concluded about the measurand. As new information becomes available the uncertainty statements must be updated to reflect the current state of knowledge.

Similarly, what constitutes a reasonable uncertainty statement must be continually updated with new information. Figure 1 displays a recent comparison of gauge block measurements and their associated expanded ($k = 2$) uncertainties [2]. In Figure 1(a) the lab denoted by the arrow has a measured value that is exactly equal to the reference value, yet has the largest measurement uncertainty of all the participants. Initially such an uncertainty statement might be considered to be far too conservative. However, on a second gauge block shown in Figure 1(b) the same laboratory has a measurement uncertainty statement that just includes the reference value. In consideration of these two results the uncertainty statement might now be considered quite reasonable. A third result, shown in Figure 1(c), has the reference value far from being included in the laboratory's uncertainty statement. Consequently, given this additional information it might be reasonable to conclude that the participant is too optimistic in their uncertainty evaluation since one in three uncertainty statements do not include the reference value.

This illustrates several issues concerning uncertainty statements. (1) It is possible to have (unknowingly) a small error yet a large uncertainty statement. This might occur when little information is available on some influence quantity resulting in assigning a relatively large standard uncertainty. Hence a large uncertainty relative to the observed errors could be a reasonable statement since it satisfies the criteria of "encompasses a large fraction of the values that can reasonably be attributed to the measurand." Conversely, if small measurement errors are repeatedly observed, for example when using calibrated check standards to evaluate the measurement process, this new information should be used to reexamine and reduce the measurement uncertainty. (2) An uncertainty evaluation should include only a "single significant figure," that is, agreement between experts at the 10 % level should be considered very good. (3) It is relatively easy to invalidate an uncertainty statement compared to validating one, by examining measurement errors. For example, if three out of five measurement errors (determined by measuring calibrated artifacts) lie outside of the uncertainty statement, this alone would be strong evidence that such an uncertainty statement is invalid, i.e., it does not encompass a large fraction of the reasonable values that can be attributed to the measurand.

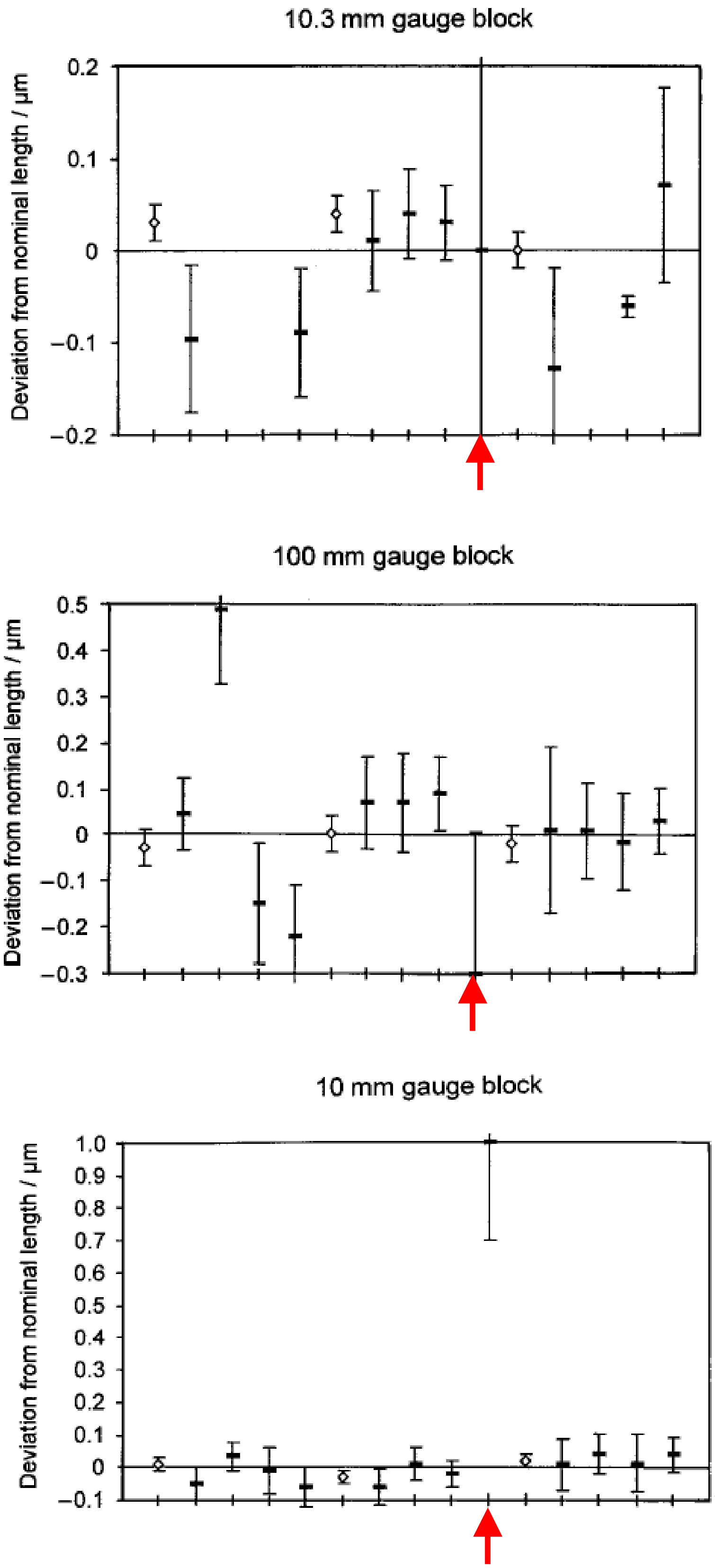


Figure 1
Results from a recent international gauge block comparison. All uncertainty bars are for a coverage factor of $k = 2$

From [2]

(a)

(b)

(c)

One aspect of uncertainty evaluation that makes it a considerably more difficult task than classical (frequentist) statistics would suggest is the issue of systematic errors. Such errors do not reveal themselves in the measurement data as observed variation. Only upon measurement of a calibrated artifact, that embodies the measurand of interest, will systematic errors become apparent. In many cases such calibrated artifacts are not available. Figure 2 illustrates this situation in the measurement of the Newtonian gravitational constant *G,* showing the reference value of *G* and its associated combined standard (k = 1) uncertainty from the 1986 CODATA report [3]. In 1995 a very credible laboratory produced a result that differed from the reference value by an amount that was 40 times the root sum-of-squares (RSS) of its standard uncertainty and the standard uncertainty of the 1986 CODATA value. Despite an extensive review of their uncertainty evaluation, no additional contributors could be discovered. Clearly, it must have been an agonizing decision for the researchers to go to press with these results. Recent measurements have not confirmed their extraordinary value, but rather, are in close agreement with the 1986 reference value.

Despite several reviews of the experiment by numerous experts, the source of the systematic error (bias) remains a mystery. The experiment employed a nontraditional (for this type of measurement) apparatus and hence the experimenters did not have the benefit of an established pool of experience from which todraw. Although the 1998 CODATA value of *G* was kept at the 1986 reference value, the discrepancy in the published values resulted in the 1998 combined standard uncertainty value being increased 12 times relative to the 1986 uncertainty. This example illustrates two important issues. (1) It can be extremely difficult to detect some systematic errors and appropriately account for them in the uncertainty statement. (Corollary, take full advantage of calibrated artifacts if they are available, for they embody the true value.) (2) In some unusual situations, additional information can result in an *increase* in the uncertainty of a quantity since the new information may reveal effects that have been previously overlooked.

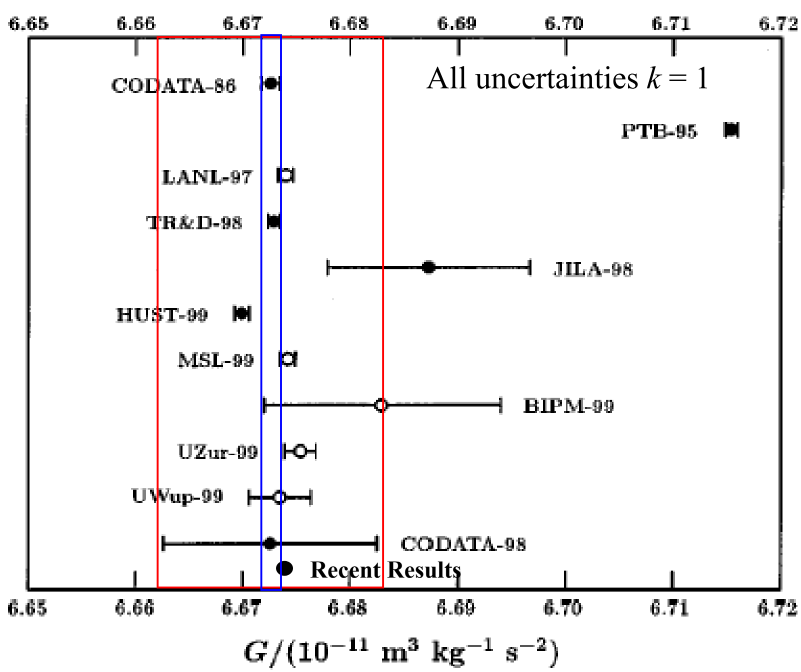


Figure 2
Measurements of the Newtonian gravitational constant. All uncertainty bars are combined standard uncertainties (k = 1). The vertical lines show the standard uncertainty of the 1986 and 1998 CODATA values.

From [3]

Systematic errors are often present in dimensional metrology and must be accounted for in the uncertainty budget. Unfortunately, as shown in the previous example, it is often difficult to recognize and appropriately model these errors. Frequently, such errors arise due to differences between the measurand and the quantity that the measurement system realizes. Rarely does a measurement system realize exactly the measurand in its results. Rather, there are a series of corrections, due to systematic effects, that must be applied to bring the realized quantity into correspondence with the measurand. A few well known (and hence well characterized) effects include probe penetration into an object's surface when the measurand specifies a zero penetration result, the thermal expansion of an object due to a nonstandard temperature which must be corrected back to its length at 20 ºC, and the use of finite sampling over a surface when the measurand is defined as the entire (infinite density sampling) surface.

Figure 3(a) shows the results of an international ring gauge comparison [4]. The measurand is the two-point diameter at a specified orientation and mid-height (relative to the face of the gauge) of the ring gauge. In this particular comparison, the gauges ranged in diameter from 3 mm to 90 mm. Since these are typical diameters for this type of artifact, all of the participants used well-characterized contact probing techniques. In Figure 3(a) the results for a 3 mm ring gauge show that the expanded uncertainties overlap for nearly all of the participants, indicating good agreement of the results. Figure 3(b) shows a different international comparison (but involving most of the same participants) of small ring gauges [5]. Due to the small diameter, some laboratories elected to use optical measurement systems; in particular, one method was limited to measuring the diameter near the face of the ring gauge instead of the mid-height (a few hundred micrometers below the surface of the gauge). As it turned out, the diameter measurements made at the surface (method C in Figure 3 (b)) were systematically large, due to edge effects [5].

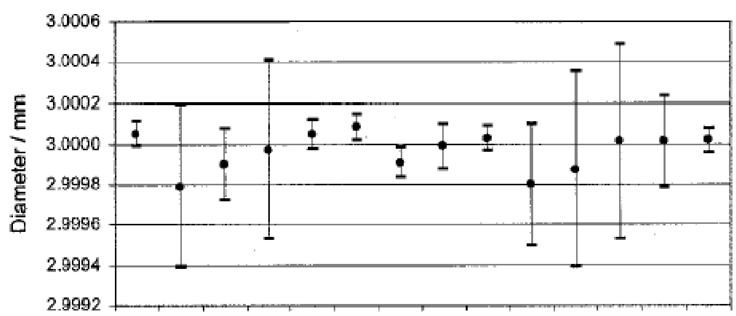
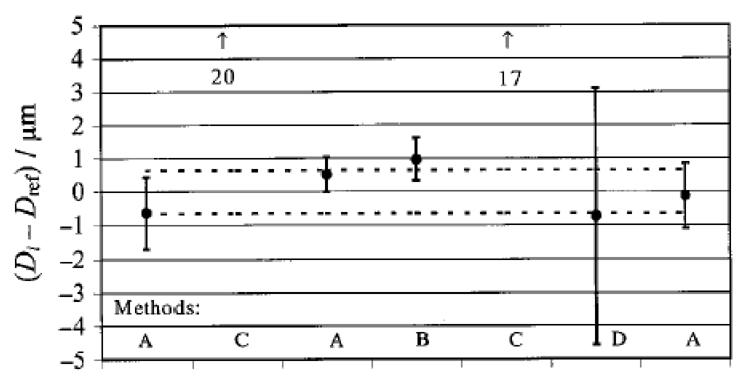


Figure 3
(a) An international comparison of a 3 mm diameter ring gauge; all participants used contact gauging and the results are in good agreement. [4]



(b) Another international comparison of a small, 0.3 mm diameter ring gauge. Note that labs using method C show large systematic deviations from the reference value. [5]

The issue of geometrical imperfections, as manifested by artifact form error, leading to less reliable uncertainty statements has been observed in a number of comparisons. Even in comparisons where the measurand is the form error of the artifact, it typically leads to a poor estimation of uncertainty. Jusko *et. al*. [6] reported in a comparison of artifacts measured for form error that most of the participants had overlapping uncertainty intervals when the artifacts had submicrometer form error, but when the artifact geometry was less than ideal, the uncertainty intervals did not overlap for over 50 % of the results. Similar problems were noted by Hansen *et. al*. [7] involving a comparison of artifacts measured on CMMs. In that study it was found that when a geometrically simple measurand was measured, most participants successfully included the reference value in their uncertainty interval; see Figure 4 (a). In contrast, when the measurand was more complex geometrically, such as the coaxiality between two bores, nearly half the participants had difficulty producing reasonable uncertainty statements; see Figure 4(b).

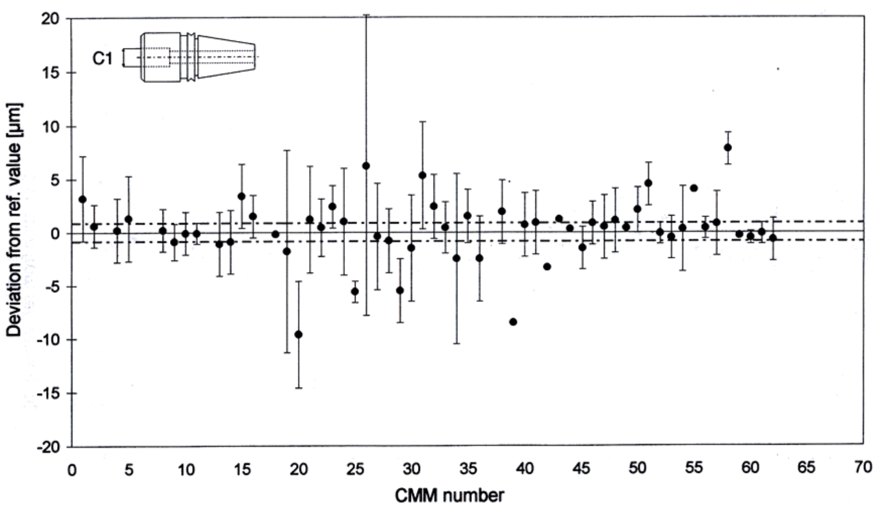
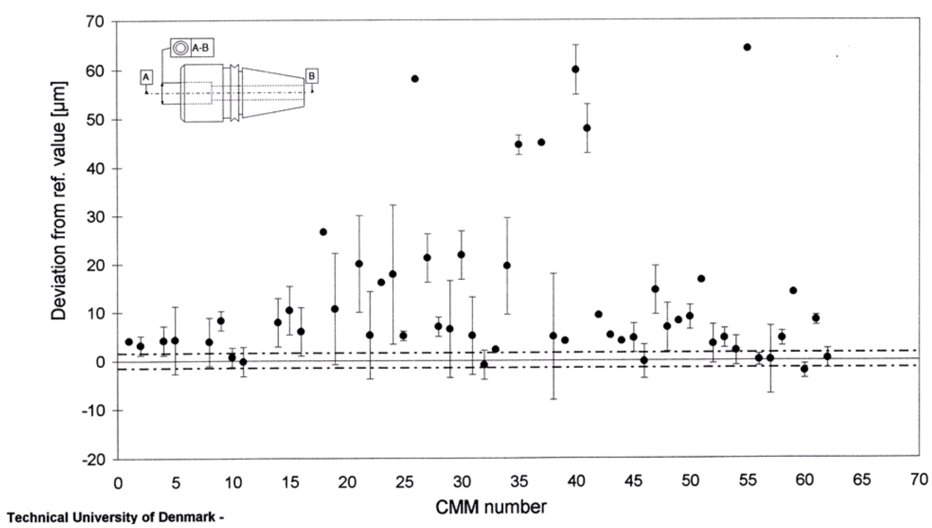


Figure 4; from [7]

(a) A comparison of the bore diameter of a tool holder. Of the 48 participants that submitted expanded (k = 2) uncertainty statements, 43 (almost 90 %) included the reference value in their interval. This is close to the expected 95 % anticipated by the GUM



(b) The same comparison measuring a different feature on the tool holder. Of the 40 participants that submitted expanded (k = 2) uncertainty statements for the coaxiality of two bores, only 21 (52.5 %) included the reference value in their interval.

Numerous laboratory intercomparisons have shown that metrologists can reliably construct reasonable expanded uncertainty statements under well-characterized, i.e., “standardized” measurement conditions, even when the measurement equipment is complex and the uncertainty statement is very accurate (e.g., multi-color interferometry with less than 10 nm combined standard uncertainty [8]). Usually these measurements involve geometrically simple measurands, nearly ideal artifacts (both in form, material, and nominal size), well-studied measurement systems, and standard environmental conditions. While it may seem obvious that that magnitude of measurement errors will increase as a measurement deviates from these standardized conditions, it is not obvious what will happen with regard to the uncertainty statement. One can easily imagine a world (of pessimists) where the uncertainty statement greatly increases in magnitude to reflect the metrologist’s concern over nonstandard measurements. This does not seem to be the case today. The fact that metrolgists do not sufficiently enlarge their uncertainty statements for non-standardized measurements is more an attribute of human nature than metrology. Some of the factors that seem to affect the validity of uncertainty statements are shown in the table below.

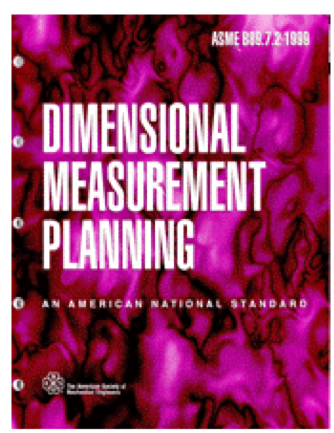
Table of Factors Affecting the Likelihood of the Validity of Uncertainty Statements

Influence Factor:	Increases the likelihood of the uncertainty interval’s validity:	Decreases the likelihood of uncertainty interval’s validity:
Geometrical Complexity	Simple, e.g. gauge blocks	Complex, e.g. concentricity
Surface Perfection	Perfect, e.g. lapped	Poor, e.g. large out-of-roundness
Material Properties	Widely used, e.g. gauge steel	Infrequently used
Knowledge Pool of Measurement Technique	Extensive, e.g. interferometry	Little, novel techniques
Personal Experience	Substantial	Little
Uncertainty Templates	Extensive, e.g. CCL comparisons	None

Resources for Evaluating Measurement Uncertainty

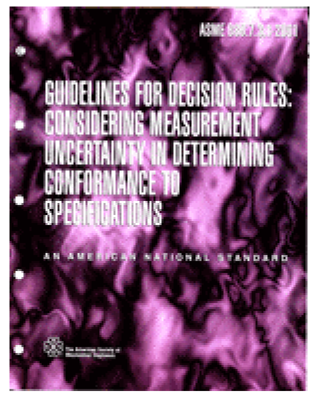
In an effort to increase the available knowledge base and provide standardized techniques for uncertainty evaluations, numerous national and international standards are under development. In the US the ASME B89.7 committee was formed to provide some guidance on these topics, with emphasis on industrial (shop floor) metrology issues. The remainder of this paper discusses some of the current and future work of this and other standards committees.

B89.7.2 (1999)



The B89.7.2 Standard “Dimensional Measurement Planning” [9] is the first published document from the B89.7 series. As the name implies, this standard is an overview of the entire measurement process. The formal part of the standard is just a brief three pages long, essentially a list of factors to consider when developing measurement plans. The bulk of the document consists of appendices that include a worked example and supporting information. B89.7.2 is a high level document. It starts by asking what measurements need to be performed, why they are being performed (e.g. process control), and what fraction (lot sampling) of the workpieces needs to be measured. Consideration is then given to selection of the measuring equipment, development of the measurement strategy, and calculation of the measurement uncertainty. Issues such as the probability of rejecting a workpiece that is within specifications or accepting one that is out of specification are also presented. Additional factors such as operator skill, the location of the measurements, measurement cycle time, and some documentary requirements are also discussed. Do not expect this standard to provide detailed information on how to make measurements or how to calculate measurement uncertainty; these are left for other more specific documents. B89.7.2 is a unique standard from the perspective of addressing the entire dimensional measurement process yielding a concise list of requirements for measurement planning.

B89.7.3.1 (2001) & ISO 14253-1 (1998)



Published in the March of 2002, the B89.7.3.1 [10] document "Guidelines for Decision Rules: Considering Measurement Uncertainty in Determining Conformance to Specifications" specifically addresses the issue of applying measurement uncertainty in industrial settings. A decision rule is a prescription for the acceptance or rejection of products based on the measurement result of a characteristic of the product, the permissible variation associated with that characteristic, and the uncertainty of the measurement result. For workpieces the permissible variation is commonly called the tolerance; for instruments it is often given by the specification limits or a maximum permissible error (MPE). We will adopt the terminology of ISO 14253-1 and refer to the permitted variation of a product's characteristic as the specification zone.

Some measurements, particularly at NMIs, state a description of the measurement, its result, and its uncertainty; decision rules are not involved since there are no product specifications. However, most industrial measurements are performed to determine if a product is in accordance with some specification, e.g., if a workpiece is within its specified tolerance. In this situation, the measurement value is usually used in a binary decision that the product is acceptable or not acceptable. This general class of problems, determining if a measurement result yields an acceptable product when clouded by measurement uncertainty, is not addressed by the GUM and represents an important economic (and potentially conflict prone) application of measurement uncertainty.

The concept of a decision rule has a long history and over the years it has developed many variations including "gauge maker's rule," "test accuracy ratio" (TAR), "test uncertainty ratio" (TUR), "four-to-one rule," "gauging ratio," "guard bands," "gauging limit," and many more. Most of these terms were defined before the development of the GUM and hence concepts such as "accuracy" or "uncertainty" were nebulously defined.

The goals of the B89.7.3.1 document are to establish a set of requirements for a decision rule, define a terminology that allows unambiguous communication of what decision rule is being used, and provide some well-documented decision rules that can be referenced.

Briefly stated, a decision rule must meet four conditions: (1) A decision rule must have a well-documented method of determining the location of the acceptance, rejection, and any transition zones (transition zones are optionally defined regions between acceptance or rejection; see Figure 8). (2) Each zone of a decision rule must correspond to a documented decision that will be implemented should the result of a measurement lie in that zone. While this is implicit for the acceptance and rejection zones by definition, any transition zones must have their corresponding decision outcomes documented. (3) A decision rule must state the procedure for addressing repeated measurements of the same characteristic on the same workpiece or instrument. (4) A decision rule must state the procedure regarding data rejection, that is, rejection of “outliers.”

The most common form of acceptance and rejection used in industry is the descendant of the “four-to-one rule” given in MIL-STD 45662A. In the new terminology this is called *simple 4:1 acceptance.* This requires that the magnitude of the expanded (k = 2) measurement uncertainty interval (± U) is no larger than the 1/4 of the specification zone (hence the expanded uncertainty, U, is to be no larger than 1/8 of the specification zone), and that product is accepted if the measurement result lies in the specification zone and rejected otherwise (see Figure 5).

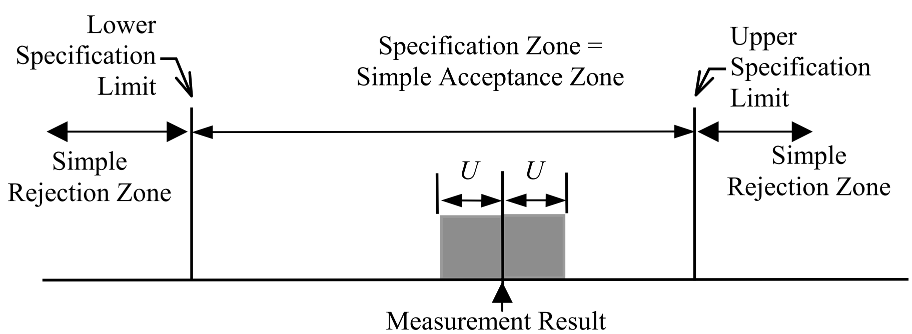


Figure 5. An example of a simple 4:1 acceptance decision rule. The measurement uncertainty interval is of width 2 U, where U is the expanded uncertainty, and the uncertainty interval is no larger than one-fourth the product’s specification zone. The measurement value shown results in product acceptance.

The simple acceptance decision rule, while straightforward, has difficulties near the specification zone limits. Due to measurement uncertainty, a product with a measurement result just inside the specification limit may actually be nonconforming. If accepting nonconforming parts has a large negative economic impact, then implementing guard banding is preferred. Guard banding is a technique that can produce a *stringent acceptance zone* that is smaller than the specification zone due to the guard bands (see Figure 6).

The size of the guard band, g, is expressed as a percentage of the expanded uncertainty, e.g. a 100 % guard band has a magnitude equal to the expanded uncertainty. Establishing the magnitude of a guard band is a business decision and is based on economics, whereas evaluating the measurement uncertainty, U, is a technical activity that depends on the measurement process.

Descriptors such as “stringent” and “relaxed,” used in describing conformance and non-conformance, have been carefully chosen. For example, stringent acceptance implies both a *decrease* in the acceptance zone width and an *increase* in confidence that a measurement result in this zone is associated with an in-specification product. Similarly stringent rejection results in a decreased size of the rejection zone while increasing the confidence that a measurement result in this zone is associated with an out-of-specification product. The converse situation applies to relaxed acceptance and rejection.

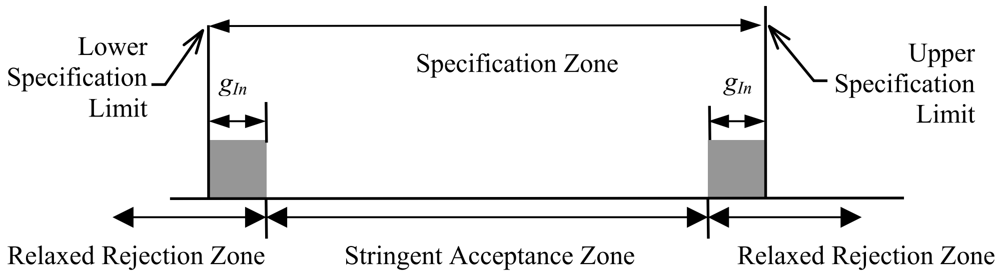


Figure 6. Stringent acceptance and relaxed rejection using symmetric two-sided guard banding. Products are accepted if the measurement result is within the acceptance zone.

If the product costs are very high, and the cost of accepting a nonconforming product is low, then *relaxed acceptance* may be preferred, see Figure 7. Relaxed acceptance allows an acceptance zone that is larger than the specification zone. This is a useful rule when a design requirement has resulted in a specification zone comparable to the state-of-the-art measurement uncertainty and hence even simple acceptance will result in a large number of conforming products being rejected.

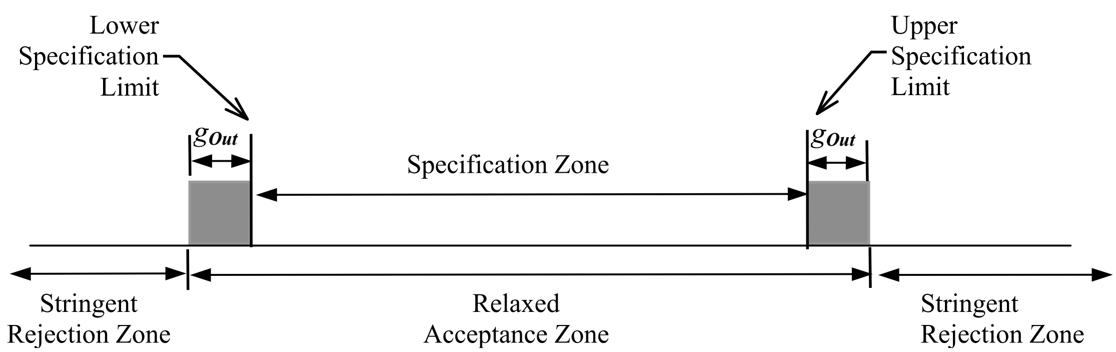


Figure 7: Relaxed acceptance and stringent rejection using symmetric two-sided guard banding. Products are accepted if the measurement result is within the acceptance zone.

Other decision rules are also possible. Figure 8 shows a situation with stringent acceptance, simple rejection, and a transition zone where the product is likely to be in conformance but the confidence of this statement is lower than that for measurement results in the stringent acceptance zone. The outcome of a measurement result in the transition zone could be, for example, selling the product at a reduced price. Ultimately the selection of a particular decision rule is a business decision that is economically driven; some of the factors to be considered are outlined in appendices of the B89.7.3.1 document.

The B89.7.3.1 document is similar to the ISO standard 14253-1 [11]. The ISO 14253-1 document focuses on the case of using stringent acceptance with a 100 % guard band for the supplier of a product and stringent rejection with a 100 % guard band for a customer seeking to reject a product. The B89.7.3 working group believes that the selection of a decision rule is a business decision, and the flexibility of having a continuum of rules ranging from stringent to relaxed acceptance or rejection is needed in order to satisfy a broad range of industries. (In B89.7.3.1, guard bands can have any percentage of the expanded uncertainty appropriate for the economics of that measurement, and it provides the terminology to communicate the type of decision rule.) Additionally, The B89.7.3.1 document establishes the requirements of a decision rule such as repeated measurements, data rejection, and documented decision outcomes, all of which are not addressed in the ISO standard.

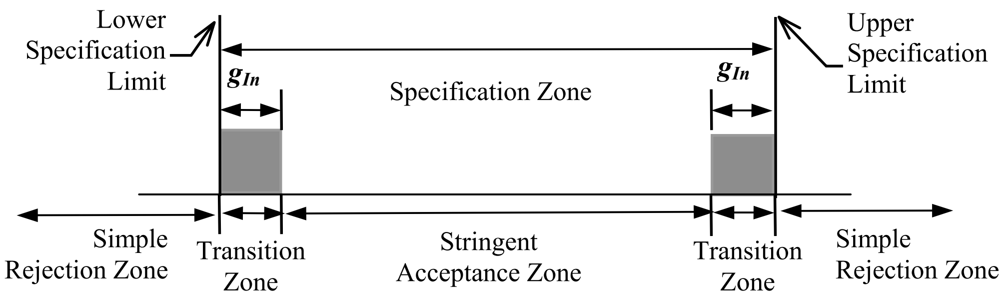


Figure 8. Stringent acceptance, simple rejection and a transition zone example using symmetric two-sided guard banding. Products are accepted if the measurement result is within the acceptance zone, rejected if in the rejection zone, and subject to a different rule in the transition zone.

ISO 14253-2 (1999)

The "Guide to the Estimation of Uncertainty in GPS Measurement in Calibration of Measuring Equipment and in Product Verification" [12] is a technical report about evaluating measurement uncertainty. The strengths of this document include a list of sources of uncertainty common to dimensional measurements and some information on evaluating these uncertainty sources.

Additionally the report describes the "Procedure for Uncertainty MAnagement" (PUMA) method of approaching uncertainty budgets. This method suggests that the first iteration of an uncertainty budget should be a rough estimate that overestimates the uncertainty by assigning relatively large values to the uncertainty contributors and lumping many uncertainty sources together as a single contributor (i.e. input quantity). The resulting uncertainty is compared against the required application. If the evaluated uncertainty is too large, e.g. the corresponding decision rule rejects too many products, then a second iteration of the uncertainty budget is performed. The second iteration may involve both a reduction in the magnitude of various uncertainty contributors, e.g. by more careful investigation into the measurement system, and perhaps a more detailed uncertainty budget. The process is repeated until the evaluated uncertainty is sufficient for the application or the uncertainty does not decrease with additional iterations, indicating another measurement method is required.

Unfortunately, 14253-2 is also filled with quite a lot of jargon and unnecessary terminology such as "true uncertainty," "conventional true uncertainty," "approximated uncertainty," and the GUM advises against using such terms. There are also philosophical differences between the GUM and the PUMA method as the GUM clearly cautions against knowingly overestimating the measurement uncertainty; nevertheless, the PUMA method is useful for industrial settings where the uncertainty of a workpiece is unlikely to be propagated into another measurement system.

B89.7.3.3 (2002) & ISO 14253-3 (2002)

B89.7.3.3 "Guidelines For Assessing the Reliability of Dimensional Measurement Uncertainty Statements" [13] is a report designed to allow parties avoid potential, or resolve actual, disagreements over the magnitude of a stated measurement uncertainty, particularly when that uncertainty is part of determining the conformance of a product to dimensional specification. With significant economic interests at stake, it is not surprising that customers and suppliers might disagree over the magnitude of the measurement uncertainty statement. Applying these guidelines can assist businesses in avoiding disagreements about measurement uncertainty statements between customers and suppliers and in resolving such disagreements should they occur. Disagreements over uncertainty statements involving a single measurement system and multiple measurement systems (each having its own uncertainty statement) are considered. Guidance is provided for examining uncertainty budgets as the primary method of assessing their reliability. Additionally, resolution by direct measurement of the measurand is also discussed. While the document was initially designed for resolving disagreements over measurement uncertainty, the report discusses many factors of formulating uncertainty budgets that will be useful to anyone responsible for this task.

The ISO 14253-3 [14] document, like the B89.7.3, is concerned with resolving disagreements between two parties over a dispute of a measurement uncertainty statement. The ISO document tends toward a more formal flowchart approach than its US counterpart, which is more tutorial in focus. The main goal of ISO 14253-3 is achieving agreement between the parties whereas B89.7.3 emphasizes the metrological issues seeking to achieve agreement through education.

Future B89.7 documents

The B89.7 committee has several ongoing work items. The B89.7.4 working group is near completing a draft of a report that addresses the quantitative risk assessment of decision rules. This document provides mathematical guidance for determining the fraction of conforming parts rejected and nonconforming parts accepted for various decision rules under various assumptions of the production and measurement uncertainty probability distributions. This report is anticipated to be available in 2004.

Also under development is B89.7.3.2, which examines a simplified GUM approach to measurement uncertainty. Finally, the B89.7.8 working group is considering issues associated with measurement traceability and how to achieve it. This document will discuss the differences between traceability and measurement uncertainty and the steps needed to provide the required documentation.

Other Recent Documents Available On Line

Fortunately, several documents on measurement uncertainty are available for free and can be downloaded from the Internet. A brief description of each document is presented and a URL address given.

NIST "Technical Note 1297" [15] is sometimes called a summary of the GUM; it contains the basic definitions and method. While the GUM is highly recommended, TN 1297 is available at no cost and hence could be freely distributed within an organization.
URL: http://physics.nist.gov/Document/tn1297.pdf

An online introduction to the GUM and the SI system of units is available on NIST's web site.
URL: http://physics.nist.gov/cuu/Uncertainty/index.html

"A Careful Consideration of the Calibration Concept" [16] is primarily focused on uncertainty issues related to calibrations, i.e. measurements that result in the issuing of certificates describing the accuracy of the measurement. The paper does contain an introduction useful for all metrologists in defining the measurement under consideration, i.e. the measurand. This often-overlooked factor can result in protracted arguments between suppliers and customers, for example a supplier might measure the diameter of a bore as acceptable using a plug gauge while the customer rejects this feature using a least-squares fit on a coordinate measuring machine. No amount of improvement in the accuracy of these measuring methods will resolve this discrepancy as two fundamentally different measurands are under inspection. The document also has an extensive appendix that is a tutorial on basic issues of uncertainty such as the distinction between measurement uncertainty and error.
URL: http://nvl.nist.gov/pub/nistpubs/jres/106/2/j62phi.pdf

"Uncertainty and Dimensional Calibrations" [17] provides an excellent discussion of sources of uncertainty relevant to the calibration of dimensional artifacts and gauges. A generic uncertainty budget is presented and nine examples, including gauge blocks, ring gauges, optical flats, and sieves are worked out in detail. While the intended audience is gauge calibration laboratories, all metrologists will benefit from the clear presentation and application of the GUM to several different measurement situations.
URL: http://nvl.nist.gov/pub/nistpubs/jres/102/6/j26doi.pdf

For the advanced measurement uncertainty expert the following papers may be of interest.

"The Calculation of Measurement Uncertainty using Prior Information" [18] discusses a Bayesian inference approach to including prior information about the value of the measurand in the calculation of measurement uncertainty;
URL: http://nvl.nist.gov/pub/nistpubs/jres/103/6/j36phi.pdf

"Guidelines for Expressing the Uncertainty of Measurement Results Containing Uncorrected Bias" extends the GUM to the case of including uncorrected systematic errors in an expanded measurement uncertainty statement.
URL: http://nvl.nist.gov/pub/nistpubs/jres/102/5/j25phi.pdf

"A Distribution-Independent Bound of the Level of Confidence in the Result of a Measurement" [19] discusses the relationship between the coverage factor in the expanded uncertainty and the corresponding level of confidence.
URL: http://nvl.nist.gov/pub/nistpubs/jres/102/5/j25est.pdf

Finally, a good source of additional publications on measurement uncertainty can be found at the Bureau International des Poids et Measures (BIPM) website of the Joint Committees for Guides in Metrology.
URL: http://www.bipm.org/CC/documents/JCGM/bibliography_on_uncertainty.html

Summary

Globalization, international quality standards, and economic factors are rapidly driving interest in measurement uncertainty. There is good capability to produce reasonable uncertainty statements for geometrically simple measurands and correspondingly geometrically perfect artifacts. However, this capability rapidly degrades for many metrologists as the measurand becomes more complex, or the artifacts geometrically imperfect. In both the US and at the ISO level, committees are actively working to supply information to industry that will address this issue.

Acknowledgements

The author would like to thank Dr. Tyler Estler, Mr. Bruce Borchardt, and Dr. Craig Shakarji, for their review of the manuscript. This work was supported in part by the Air Force's Coordinated Calibration Group, and the Manufacturing Engineering Laboratory of NIST.

References

1 International Organization for Standardization, "Guide to the Expression of Uncertainty in Measurement," Geneva, Switzerland, 1993. (corrected and reprinted 1995). This document is also available as a U.S. National Standard: NCSL Z540-2-1997.
2 "International Comparison Final Report", *Metrologia*, **38**, 273-275 (2001).
3 "CODATA Recommended Values," P.J. Mohr and B.N. Taylor, *Rev. Mod. Phys*., Vol. 72, No. 2, April 2000.
4 "Key and International Comparison, EUROMET key comparison: Cylindrical diameter standards," R. Thalmann, *Metrologia*, 37, 253-260 (2000).
5 "International Comparison, EUROMET comparison: diameter of small ring gauges," M. Neugebauer and F. Lüdicke, *Metrologia*, **38**, 259-267 (2001).
6 "Results of the CIRP Form Intercomparision 1996-1998," O. Juslo, J.G. Salsbury, and H. Kunzmann, Annals of the CIRP, **48,** 413-415 (1999).
7 "Nordic audit of coordinate measuring machines. Final report." H.N. Hansen and L. DeChiffre, Dept. of Manufacturing Engineering Technical University of Denmark, Lyngby, Denmark, 1996.
8 "CCL Key Comparison: calibration of gauge blocks by interferometry," R. Thalmann, *Metrologia*, **39**, 165-177 (2002).
9 ASME B89.7.2-1999 "Dimensional Measurement Planning," available from the American Society of Mechanical Engineers. www.asme.org.
10 ASME B89.7.3.1-2001 "Guidelines for Decision Rules: Considering Measurement Uncertainty in Determining Conformance to Specifications," available from the American Society of Mechanical Engineers. www.asme.org.
11 International Standard, "Geometrical Product Specifications (GPS) - Inspection by measurement of workpieces and measuring instruments -- Part 1: Decision rules for providing conformance or non-conformance with specification," ISO 14253-1 (1998).

12 International Standard, "Geometrical Product Specifications (GPS) - Inspection by measurement of workpieces and measuring instruments -- Part 2: Guide to the Estimation of Uncertainty in GPS Measurement in Calibration of Measuring Equipment and in Product Verification," ISO 14253-2 (1999).
13 ASME B89.7.3.3-2002 "Guidelines For Assessing the Reliability of Dimensional Measurement Uncertainty Statements," available from the American Society of Mechanical Engineers; www.asme.org.
14 International Standard, "Geometrical Product Specifications (GPS) - Inspection by measurement of workpieces and measuring instruments -- Part 3: Guidelines to achieving agreements on measurement uncertainty statements," ISO 14253-3 (2002).
15 "Guidelines for Evaluation and Expressing the Uncertainty of NIST Measurement Results," B.N. Taylor and C.E. Kuyatt, NIST Technical Note 1297, 1994 Edition, National Institute of Standards and Technology, Gaithersburg, MD 20899, 1994.
16 "A Careful Consideration of the Calibration Concept," S.D. Phillips, W.T. Estler, T. Doiron, K.R. Eberhardt, M.S. Levenson", *Journal of Research of NIST*, **106**, 1 – 9 (2001).
17 "Uncertainty and Dimensional Calibrations," T. Doiron and J. Stoup, *Journal of Research of NIST*, **102**, 647-676, (1997).
18 "Calculation of Measurement Uncertainty Using Prior Information," S.D. Phillips, and W.T. Estler, *Journal of Research of NIST*, **103**, 625-632 (1998).
19 "A Distribution-Independent Bound of the Level of Confidence in the Result of a Measurement," W.T. Estler, *Journal of Research of NIST*, **102**, 587-588 (1997).